

# VOLUME VS. RANK OF LATTICES

T. GELANDER

ABSTRACT. We show the rank (i.e. minimal size of a generating set) of lattices cannot grow faster than the volume.

## 1. THE MAIN RESULT AND SOME CONSEQUENCES

For a group  $\Gamma$  we denote by  $d(\Gamma)$  the minimal cardinality of a generating set. When considering the class of finite index subgroups  $\Gamma$  in some given finitely generated group  $\Delta$ , it is easy to show that  $d(\Gamma)$  is at most  $(d(\Delta) - 1)[\Delta : \Gamma] + 1$  and in particular bounded linearly by the index. We prove the analog statement when  $\Delta$  is replaced by a connected semisimple Lie group  $G$ .

**Theorem 1.1.** *Let  $G$  be a connected semisimple Lie group without compact factors. Then there is an effectively computable constant  $C = C(G)$  such that*

$$d(\Gamma) \leq C \text{vol}(G/\Gamma),$$

*for every irreducible lattice  $\Gamma \leq G$ .*

Theorem 1.1 for rank one groups was proved in [BGLS, Section 2]. For torsion free lattices Theorem 1.1 was proved in [Ge] where also bounds on the number of *short* relations were given. The proof of the general case encounters several difficulties which do not appear in rank one and in the torsion free case.

Theorem 1.1 implies in particular the well known but nontrivial fact that every lattice is finitely generated. Originally the finiteness of  $d(\Gamma)$  was proved case by case by many different authors, notably is the work of Garland and Raghunathan [GR] for  $\text{rank}(G) = 1$  and of Kazhdan [K] for the case where all the factors of  $G$  have  $\text{rank} \geq 2$ . In [BGS] a geometric proof was given under no assumption on  $G$  but assuming that  $\Gamma$  is torsion free, i.e. that  $\Gamma \backslash G/K$  is a manifold. Furthermore, in [BGS] an analog result was proved for more general manifolds of nonpositive curvature which are not necessarily locally symmetric. Our proof is inspired by [BGS], but holds under no assumption on  $G$  or on  $\Gamma$ .

---

Note in particular that our proof does not rely on Kazhdan's property (T).

Note also that Theorem 1.1 implies the classical Kazhdan–Margulis theorem:

**Corollary 1.2** (Kazhdan–Margulis Theorem [KM]). *For any semisimple Lie group without compact factors  $G$ , there is a positive lower bound on the covolume of lattices.*

*Proof.* Indeed,

$$\mathrm{vol}(G/\Gamma) \geq \frac{d(\Gamma)}{C} \geq \frac{2}{C},$$

for every lattice  $\Gamma \leq G$ . □

Our proof also gives the stronger statement of Kazhdan–Margulis’ theorem that there is an identity neighborhood in  $G$  which intersects trivially a conjugate of every lattice (see Remark 2.12).

In general, the linear estimate in Theorem 1.1 cannot be improved. For instance, for every  $n \geq 2$  the group  $G = \mathrm{SO}(n, 1)$  admits a lattice  $\Gamma$  which projects onto the free group  $F_2$  (see [Lu1]), hence taking  $\Gamma_k$  in  $\Gamma$  to be the pre-image of an index  $k$  subgroup of  $F_2$ , we get  $d(\Gamma_k) \geq k + 1$  while  $\mathrm{vol}(G/\Gamma_k) = k \cdot \mathrm{vol}(G/\Gamma)$ . On the other hand, in general one cannot give a lower bound on  $d(\Gamma)$  which tends to  $\infty$  with  $\mathrm{vol}(G/\Gamma)$ . Moreover, by [SV], when  $\mathrm{rank}_{\mathbb{R}}(G) \geq 2$ , every nonuniform lattice in  $G$  admits a 3-generated finite index subgroup.

Another immediate consequence of Theorem 1.1 is that the first Betti number grows at most linearly with the covolume:

**Corollary 1.3.** *The first Betti number  $b_1(\Gamma, A)$  of a lattice  $\Gamma \leq G$ , with respect to an arbitrary ring  $A$ , is at most  $C\mathrm{vol}(G/\Gamma)$ .*

*Proof.*

$$b_1(\Gamma, A) \leq d(\Gamma) \leq C\mathrm{vol}(G/\Gamma).$$

□

Corollary 1.3 extends Theorem 2 from [BGS] to locally symmetric orbifolds rather than manifolds. However, while in [BGS] it is shown that *all* the Betti numbers are bounded by the volume of the manifold, here we obtain a bound only on the first Betti number. It is likely that the higher Betti numbers of orbifolds are bounded by the volume as well. We refer to [FGT] for some lower bounds on  $b_1$  of some arithmetic lattices in  $\mathrm{SL}_2(\mathbb{C})$ . When  $G$  has property (T), the first Betti number with coefficients in  $\mathbb{Z}$ ,  $b_1(\Gamma, \mathbb{Z})$  vanishes for every lattice  $\Gamma \leq G$ , however for rings with torsion  $b_1(\Gamma, A)$  is typically nonzero. For instance, by [Lu2], every finitely generated linear group admits a finite index subgroup which is mapped onto  $\mathbb{F}_p$  where  $p$  is an arbitrary rational prime.

As another consequence we obtain a concrete upper bound on the number  $AL_G(v)$  of conjugacy classes of arithmetic subgroups of  $G$  of covolume at most  $v$ :

**Corollary 1.4.** *Let  $G$  be a connected semisimple lie group without compact factors. Then for every  $\epsilon > 0$*

$$AL_G(v) \leq v^{(1+\epsilon)Cv},$$

for all  $v \gg 0$ .

Corollary 1.4 was proved in [BGLS] for rank one groups (see [BGLS, Theorem 1.1]). Using Theorem 1.1 instead of the rank one case given in [BGLS], Corollary 1.4 is proved by the argument given in [BGLS] also for higher rank groups. It was also shown in [BGLS] that for  $SO(n, 1)$  this estimate is sharp in the sense that  $AL_{SO(n,1)}(v) \geq v^{av}$  for some constant  $a > 0$  and  $v$  sufficiently large, and for  $G = SL(2, \mathbb{R})$  the following precise asymptotic formula was given

$$\log AL_{SL(2, \mathbb{R})}(v) \sim \frac{v}{2\pi} \log v$$

where the Haar measure of  $SL(2, \mathbb{R})$  is the one corresponding to the hyperbolic measure on  $SL(2, \mathbb{R})/SO(2)$ . When  $\text{rank}(G) > 1$  the estimate given in Corollary 1.4 is somewhat weak. In fact Belolipetsky and Lubotzky [BL] have shown recently that if all lattices in  $G$  posses the Congruence Subgroup Property then

$$v^{a \log v} \leq AL_G(v) \leq v^{b \log v},$$

for some  $a, b > 0$  and all  $v \gg 0$ . It was conjectured by Serre that the CSP holds for all irreducible lattices in all higher rank semisimple Lie groups, and this conjectured was confirmed in many cases. But the general case, and in particular when  $\Gamma$  is a uniform lattice in  $SL_{n \geq 3}(\mathbb{R})$ , is unknown. Corollary 1.4 holds under no assumptions. An upper bound as in Corollary 1.4 was proved for *torsion free* lattices (which are not necessary arithmetic) in [BGLM, Ge] when  $G$  was assumed to be not locally isomorphic to  $SL(2, \mathbb{R})$ ,  $SL(2, \mathbb{C})$ ,  $SL(3, \mathbb{R})$  and to  $SL(2, \mathbb{R}) \times SL(2, \mathbb{R})$ . Note however that even when restricting to torsion free lattices, in the special cases of  $G = SL(3, \mathbb{R})$  or  $SL(2, \mathbb{R}) \times SL(2, \mathbb{R})$  no quantitative estimate was given until today.

**About the proof:** The basic idea is to deform the symmetric space  $X = G/K$  to nice connected  $\Gamma$ -invariant subset  $Y \subset X$  where the displacement of every  $\gamma \in \Gamma \setminus \{1\}$  is bounded below by a uniform constant, and deduce that  $\Gamma$  is a quotient of  $\pi_1(Y/\Gamma)$  while the topology of  $Y/\Gamma$  is bounded by its volume since it is uniformly thick. We use techniques similar to [BGS, Ge].

## 2. THE PROOF

**2.1. Some notation and background.** Let  $G$  be a connected center-free semisimple Lie group,  $K \leq G$  a maximal compact subgroup of  $G$  and  $X = G/K$  the associated Riemannian symmetric space.  $X$  equipped with the analytic Riemannian metric coming from the Killing form of  $G$  is non-positively curved, i.e. the distance function  $d : X \times X \rightarrow \mathbb{R}$  is convex.

For  $g \in G$ ,  $x \in X$  we denote by

$$d_g(x) = d(x, g \cdot x)$$

the displacement of  $g$  at  $x$ , and by

$$\text{Min}(g) := \{x \in X : d_g(x) = \inf(d_g)\}$$

the set where  $d_g$  attains its infimum. Recall that  $g$  is *semisimple* iff  $\text{Min}(g) \neq \emptyset$  (equivalently, iff  $g$  belongs to a torus in  $G$ ), and otherwise  $g$  is called *parabolic*.

A semisimple element  $g$  is *elliptic* if  $\inf d_g = 0$  in which case  $\text{Min}(g)$  is also denoted by

$$\text{Fix}(g) = \{x \in X : g \cdot x = x\},$$

and otherwise it is *hyperbolic* and admits an axis on which it translates by a positive amount and the set  $\text{Min}(g)$  consists of the union of all its axis. In either case  $\text{Min}(g)$  is a complete totally geodesic submanifold.

$G$  admits a structure of a linear real algebraic group, and an element  $g \in G$  is semisimple iff it is diagonalizable, and is *unipotent* iff all its eigenvalues are 1. These properties are independent of the chosen representation of  $G$ . Moreover, a unipotent element  $g \in G$  is parabolic with  $\inf d_g = 0$  (the converse is not true in general).

For  $t \geq 0$  we denote by  $\{d_g \leq t\}$  the sublevel set  $\{x : d_g(x) \leq t\}$ . Since  $X$  has nonpositive sectional curvature,  $\{d_g \leq t\}$  is convex.

**2.2. A co-dimension 2 condition.** Write  $c \parallel c'$  to express that two geodesic lines  $c, c'$  are parallel, and denote

$$\mathfrak{I}(c) := \bigcup \{c'(\mathbb{R}) : c' \parallel c\}$$

the union of all traces of geodesic lines in  $X$  parallel to  $c$ .

**Lemma 2.1.** *Let  $X$  be an irreducible symmetric space of noncompact type. If for some geodesic line  $c$ ,  $\mathfrak{I}(c)$  has codimension one, then  $X$  is isometric to the hyperbolic plane  $\mathcal{H}^2$ .*

*Proof.* Let  $p^+, p^- \in \partial X$  be the endpoints of  $c$  in the ideal boundary of  $X$ , and let  $P^+, P^- \leq G = \text{Isom}(X)^\circ$  be the corresponding parabolic subgroups. Observe that for  $g \in P^+$ ,  $g \cdot c(0) \in \mathfrak{I}(c)$  iff  $g \in P^-$ . Indeed, if  $g \in P^- \cap P^+$  then  $g \cdot c$  is parallel to  $c$ , and vice versa: if  $g \cdot c(0) \in \mathfrak{I}(c)$ , the geodesic

line through  $g \cdot c(0)$  parallel to  $c$  has one end  $p^+$ , and since this geodesic is determined by  $g \cdot c(0)$  and  $p^+ = g \cdot p^+$  it coincides with  $g \cdot c$ , and its other end point  $p^-$  has to be  $g \cdot p^-$ . Since  $P^+$  acts transitively on  $X$ , we deduce that

$$\dim P^+ / P^+ \cap P^- = \dim X - \dim \mathfrak{I}(c) = 1.$$

Furthermore, since  $P^+ P^-$  is open in  $G$ , being the big cell in a Bruhat decomposition, the orbit  $P^+ \cdot p^-$  is open in  $G \cdot p^- \sim G/P^-$ . It follows that  $G/P^-$  is one dimensional, hence, being compact, is homeomorphic to the circle  $S^1$ . However the only center free simple Lie group that acts nontrivially on  $S^1$  is  $\mathrm{PSL}_2(\mathbb{R})$ . Hence  $G \cong \mathrm{PSL}_2(\mathbb{R})$  and  $X \cong \mathcal{H}^2$ .  $\square$

**Corollary 2.2.** *Let  $G$  be a connected center-free semisimple Lie group without compact factors and not isometric to  $\mathrm{PSL}_2(\mathbb{R})$ , and let  $g \in G$  be an element whose projection to every factor of  $G$  is nontrivial. Then  $\mathrm{codim}_X \mathrm{Min}(g) \geq 2$ .*

*Proof.* If  $G$  has more than one factor,  $\mathrm{Min}(g) \cong \prod \mathrm{Min}(g_i)$  where  $g_i$  are the components of  $g$  so the conclusion is trivial since the co-dimension is at least the number of factors. Assume therefore that  $G$  is simple. Since  $G$  is connected, every  $g \in G$  preserves the orientation of  $X$ , hence  $\mathrm{Fix}(g)$  is of codimension  $\geq 2$  for every elliptic element  $g \in G$ . For  $g$  hyperbolic, since  $\mathrm{Min}(g) \subset \mathfrak{I}(c)$  where  $c$  is an axis of  $g$ , the result follows from Lemma 2.1  $\square$

**2.3. Defining the Morse function.** Recall the classical Margulis lemma (see [Th, Proposition 4.1.16]):

**Lemma 2.3.** *There are two constant  $\epsilon_G > 0$  and  $m_G \in \mathbb{N}$  such that if  $\Lambda < G$  is a discrete group generated by  $\{\gamma \in \Lambda : d_\gamma(x) \leq \epsilon_G\}$  for some  $x \in X$ , then  $\Lambda$  contains a normal subgroup of index  $\leq m_G$  which is contained in a connected nilpotent subgroup of  $G$ .*

Set

$$\epsilon = \frac{\epsilon_G}{2} \quad \text{and} \quad m = m_G!$$

We will also use the following:

**Lemma 2.4.** *There is an integer  $\nu = \nu(G)$  such that for every ascending sequence of length  $\nu$  of centralizers of elements  $g_1, \dots, g_\nu \in G$*

$$C_G(g_1) \subset C_G(g_2) \subset \dots \subset C_G(g_\nu),$$

*for some  $i < \nu$  we have  $C_G(g_i) = C_G(g_{i+1})$ .*

*Proof.* Considering  $G$  as an algebraic subgroup of  $\mathrm{GL}_n(\mathbb{C})$ , the centralizer

$$C_G(g) = \{h : hgh^{-1} = g\} \cap G$$

is defined by  $n^2$  quadratic polynomials and the polynomials defining  $G$ . Since  $C_G(g)$  is a group, its irreducible components are its connected components. Hence, by Bezout's theorem, the number of connected components of  $C_G(g)$  is bounded uniformly (independently of  $g$ ). Now if  $C_G(g_i)$  and  $C_G(g_{i+1})$  have the same dimension and the same number of components, they must coincide, and the lemma is proved since  $G$  is finite dimensional.  $\square$

Thus for any semisimple  $g \in G$  there is  $0 \leq j < \nu$  such that  $C_G(g^{m^j}) = C_G(g^{m^{j+1}})$ . Denote by  $j(g)$  the minimal such  $j$ . Set

$$\mu = m^\nu.$$

Let  $\Gamma$  be an irreducible lattice in  $G$ . Denote by  $M = \Gamma \backslash X$  the corresponding orbifold, and by  $\pi : X \rightarrow M$  the associated (ramified) covering map. For a subset  $Y \subset M$  let  $\tilde{Y} = \pi^{-1}(Y)$  be its preimage in  $X = \tilde{M}$ . For  $\gamma \in \Gamma \setminus \{1\}$  let

$$\bar{\gamma} = \gamma^{m^{j(\gamma)}} \text{ if } \gamma \text{ is hyperbolic, and } \bar{\gamma} = \gamma \text{ otherwise.}$$

Obviously  $d_{\bar{\gamma}}(x) \leq \mu d_\gamma(x)$ . Since for a hyperbolic element  $\text{Min}(g)$  is determined by  $C_G(g)$  we have

$$(1) \quad \text{Min}(\bar{\gamma}) = \text{Min}(\bar{\gamma}^m), \text{ for every hyperbolic element } \gamma \in \Gamma \setminus \{1\}.$$

Additionally, for  $\gamma_1, \gamma_2$  hyperbolic  $[\bar{\gamma}_1^m, \bar{\gamma}_2^m] = 1 \implies [\bar{\gamma}_1, \bar{\gamma}_2] = 1$ .

Note that for every  $\gamma_0 \in \Gamma$ , the set  $\{\gamma \in \Gamma : \bar{\gamma} = \gamma_0\}$  is finite, being compact and discrete. Let

$$\tilde{N} := \cup \{\text{min}(\bar{\gamma}) : \gamma \in \Gamma \setminus \{1\}, \inf d_{\bar{\gamma}} < \epsilon\}.$$

Clearly  $\tilde{N}$  is a  $\Gamma$ -invariant union of totally geodesic submanifolds, and since  $\Gamma$  is discrete, this union is locally finite. By Corollary 2.2, unless  $G \cong \text{PSL}_2(\mathbb{R})$ ,  $\text{codim}_X \tilde{N} \geq 2$ , implying that  $X \setminus \tilde{N}$  is connected.  $N = \pi(\tilde{N}) \subset M$  is a singular submanifold consisting of all ramified points and closed geodesics shorter than  $\epsilon$ .

Let  $f : \mathbb{R}^{>0} \rightarrow \mathbb{R}^{\geq 0}$  be a smooth function which tends to  $\infty$  at 0, strictly decreases on  $(0, \epsilon]$  and is identically 0 on  $[\epsilon, \infty)$ . For  $\gamma \in \Gamma \setminus \{1\}$  and  $x \in X \setminus \tilde{N}$  define

$$\begin{aligned} \phi_\gamma(x) = & \begin{aligned} & f(d_\gamma(x)) && \text{if } \gamma \text{ is parabolic} \\ & f(\mu d_\gamma(x)) && \text{if } \gamma \text{ is elliptic} \\ & f(d_\gamma(x) - \inf d_\gamma) && \text{if } \gamma \text{ is hyperbolic and } \inf d_\gamma < \epsilon \\ & 0 && \text{if } \gamma \text{ is hyperbolic and } \inf d_\gamma \geq \epsilon, \end{aligned} \end{aligned}$$

and set

$$\tilde{\psi}(x) := \sum_{\gamma \in \Gamma \setminus \{1\}} \phi_{\bar{\gamma}}(x).$$

There are only finitely many nonzero summands for each  $x$ , since  $\phi_{\bar{\gamma}}(x) \neq 0 \implies d_{\bar{\gamma}}(x) \leq 2\epsilon$ , and  $\Gamma$  is discrete. Thus  $\tilde{\psi}$  is a well defined smooth

function on  $X \setminus \tilde{N}$ . Note that  $\tilde{\psi}(x)$  tends to  $\infty$  as  $x$  approaches  $\tilde{N}$ . Since  $\tilde{\psi}$  is  $\Gamma$ -invariant it induces a function on  $M \setminus N$

$$\psi(x) = \tilde{\psi}(\pi^{-1}(x)).$$

For  $\delta \geq 0$  denote

$$\tilde{\psi}_{\leq \delta} = \{x \in X : \tilde{\psi}(x) \leq \delta\} \quad \text{and} \quad \psi_{\leq \delta} = \{x \in m : \psi(x) \leq \delta\} = \pi(\tilde{\psi}_{\leq \delta}).$$

Then  $d_\gamma(x) \geq \frac{\epsilon}{\mu}$  for every  $x \in \tilde{\psi}_{\leq 0}$  and  $\gamma \in \Gamma \setminus \{1\}$ . Thus, the injectivity radius in  $M$  at any point of  $\psi_{\leq 0}$  is at least  $\frac{\epsilon}{2\mu}$ .

**2.4. The main proposition.** The proof of Theorem 1.1 relies on the existence of a deformation retract from  $M \setminus N$  into an arbitrarily small neighborhood of  $\psi_{\leq 0}$ . The main part consists in showing that the gradient of  $\psi$  does not vanish outside  $\psi_{\leq 0}$  and hence defines a smooth vector field.

**Proposition 2.5.** *For  $x \in M \setminus N$ ,  $\nabla \psi(x) = 0$  iff  $\psi(x) = 0$ .*

It is more convenient to work in the universal cover where the proposition translates to: for  $x \in X \setminus \tilde{N}$ ,  $\nabla \tilde{\psi}(x) = 0$  iff  $\tilde{\psi}(x) = 0$ . We will make use of the following two lemmas:

**Lemma 2.6.** *Let  $\mathcal{C} \subset X$  be a closed convex set and let  $g$  be an isometry of  $X$  which preserves  $\mathcal{C}$ . Let  $c : [0, \infty) \rightarrow X$  be a geodesic ray satisfying*

$$c(0) = c([0, \infty)) \cap \mathcal{C} = P_{\mathcal{C}}(c(1)),$$

*where  $P_{\mathcal{C}} : X \rightarrow \mathcal{C}$  denotes the projection to the nearest point. Then  $h(t) := d_g(c(t))$  is a nondecreasing smooth convex function on  $(0, \infty)$ . In particular, if  $h(t_0) > h(0)$  then  $h'(t_0) > 0$ .*

**Lemma 2.7.** (i) *The intersection of a finite collection of nonempty sub-level sets corresponding to the displacement functions of commuting isometries of  $X$  is nonempty.*

(ii) *If  $A \leq G$  is an abelian subgroup consisting of semisimple elements. Then*

$$\bigcap_{g \in A} \text{Min}(g) \neq \emptyset.$$

The proofs of Lemmas 2.6 and 2.7 rely on the following simple:

**Claim 2.8.** Let  $g$  be an isometry of  $X$ , let  $\mathcal{C} \subset X$  be a closed convex  $g$ -invariant set and let  $x \in X$ . Then  $d_g(x) \geq d_g(P_{\mathcal{C}}(x))$ .

*Proof of 2.8.* Since  $\mathcal{C}$  is  $g$ -invariant,  $P_{\mathcal{C}}(g \cdot x) = g \cdot P_{\mathcal{C}}(x)$ . Since  $P_{\mathcal{C}}$  is 1-Lipschitz,

$$d_g(x) = d(x, g \cdot x) \geq d(P_{\mathcal{C}}(x), P_{\mathcal{C}}(g \cdot x)) = d(P_{\mathcal{C}}(x), g \cdot P_{\mathcal{C}}(x)) = d_g(P_{\mathcal{C}}(x)).$$

□

*Proof of Lemma 2.6.* Since  $X$  has nonpositive curvature, the distance between the geodesic lines  $c(t)$  and  $g \cdot c(t)$ ,  $h(t) = d(c(t), g \cdot c(t))$  is convex. By 2.8  $h(t) \geq h(0)$  for any  $t > 0$ . Indeed,

$$h(t) = d_g(c(t)) \geq d_g(P_{\mathcal{C}}(c(t))) = h(0).$$

Thus  $h'(0) \geq 0$ , and as  $h$  is convex,  $h'(t) \geq 0$  for every  $t \geq 0$ .

If  $h(t_0) > 0$  for some  $t_0$  then, by Lagrange's mean value theorem,  $h'(t) > 0$  for some  $t < t_0$  and since  $h'$  is monotonic,  $h'(t_0) > 0$ .  $\square$

*Proof of Lemma 2.7.* Note that if  $g_1$  and  $g_2$  commute, the function  $d_{g_2}$ , and hence its sublevel sets, are  $g_1$ -invariant. Indeed

$$d_{g_2}(g_1 \cdot x) = d(g_1 \cdot x, g_2 g_1 \cdot x) = d(g_1 \cdot x, g_1 g_2 \cdot x) = d(x, g_2 \cdot x) = d_{g_2}(x).$$

Suppose that we are given  $n$  commuting elements  $g_1, \dots, g_n$  and  $n$  non-negative numbers  $t_1, \dots, t_n$  such that  $\{d_{g_i} \leq t_i\}$  is nonempty for each  $i$ . By induction on  $n$  we may assume that  $\mathcal{A} = \cap_{i \geq 2} \{d_{g_i} \leq t_i\} \neq \emptyset$ . Then  $\mathcal{A}$  is closed convex and  $g_1$ -invariant. Hence by 2.8,  $d_{g_1}(P_{\mathcal{A}}(x)) \leq d_{g_1}(x)$  for every  $x$ . By choosing  $x$  with  $d_{g_1}(x) \leq t_1$  we deduce that  $\{d_{g_1} \leq t_1\} \cap \mathcal{A} \neq \emptyset$ , hence Part (i).

Now by Part (i) it follows that  $\mathcal{F}(F) = \cap_{g \in F} \text{Min}(g) \neq \emptyset$  for any finite subset  $F \subset A$ . Since those  $\mathcal{F}(F)$  are complete totally geodesic submanifolds of finite dimension with the finite intersection property, they must all intersect in a common point.  $\square$

*Proof of Proposition 2.5.* For  $x \in X$  let

$$\Sigma_x = \{\bar{\gamma} : \gamma \in \Gamma \setminus \{1\} \text{ and } \phi_{\bar{\gamma}}(x) > 0\}, \quad \Gamma_x = \langle \Sigma_x \rangle$$

and let  $\mathcal{N}_x \triangleleft \Gamma_x$  be a normal subgroup of index  $i_x \leq m_G$  in  $\Gamma_x$  which is contained in a connected nilpotent subgroup of  $G$  (see Lemma 2.3). Let  $\Sigma'_x$  be the finite set consists of all elements in  $\mathcal{N}_x$  which are expressible as words of length at most  $i_x$  in the alphabet  $\Sigma_x$ .

Let  $x \in X \setminus (\tilde{N} \cup \tilde{\psi}_{\leq 0})$ . We are going to show that  $\nabla \psi(x) \neq 0$ . We distinguish between three cases:

**Case 1:** Suppose first that  $\Gamma_x$  is finite. Then there is a common fixed point  $x_0$  for  $\Gamma_x$ . Since  $x \notin \tilde{N}$ ,  $x \notin \text{Fix}(\sigma)$  for  $\sigma \in \Sigma_x$ , and we may apply Lemma 2.6 with  $\mathcal{C} = \{x_0\}$ ,  $c : [0, \infty) \rightarrow X$  being the unit speed geodesic ray from  $x_0$  through  $x$  and  $g = \sigma$ , and conclude that  $\frac{d}{dt} d_{\sigma}(c(t)) > 0$  for any  $t > 0$ . Since  $f' < 0$  on  $(0, \epsilon)$  this implies

$$\nabla \tilde{\psi}(x) \cdot \dot{c}(t_0) = \left( \frac{d}{dt} \right)_{t_0} \tilde{\psi}(c(t)) = \sum_{\sigma \in \Sigma_x} \mu f'(\mu d_{\sigma}(x)) \left( \frac{d}{dt} \right)_{t_0} d_{\sigma}(c(t)) < 0,$$

where  $t_0 = d(x, x_0)$ . In particular  $\nabla \tilde{\psi}(x) \neq 0$ .



Before considering the two other cases note that since  $\mathcal{N}_x$  is contained in a connected nilpotent group, by Lie's theorem it is conjugated over  $\mathbb{C}$  to a subgroup of the upper triangular matrices, and hence its commutator consists of unipotent, hence parabolic, elements. In particular:

- (a) If  $\Gamma_x$  consists of semisimple elements then  $N_x$  is abelian,  
and since for nontrivial parabolic  $\gamma \in \Gamma_x$ ,  $\gamma^{i_x}$  is a nontrivial parabolic in  $N_x$ :
- (b) if  $\Gamma_x$  contains a nontrivial parabolic element then  $N_x$  contains a nontrivial central parabolic element.

**Case 2:** Suppose now that  $\Gamma_x$  is infinite and consists of semisimple elements. By (a)  $\mathcal{N}_x$  is abelian.

We claim that the set  $\Sigma_x$  contains a hyperbolic element. Since  $\Gamma_x$  is infinite, also  $\mathcal{N}_x$  is infinite, thus its generating set  $\Sigma'_x$  admits an element of infinite order  $\sigma_0$ . Thus if we suppose that  $\Sigma_x$  consists of torsion elements only, then  $\sigma_0$ , being a word of length at most  $m$  in  $\Sigma_x$ , would satisfy

$$d_{\sigma_0}(x) \leq m \max\{d_\gamma(x) : \sigma \in \Sigma_x\} \leq m \frac{\epsilon}{\mu},$$

but then, as  $\overline{\sigma_0} = \sigma_0^{m^{j(\sigma_0)}}$  for  $j(\sigma_0) \leq \nu - 1$  we would get

$$d_{\overline{\sigma_0}}(x) \leq m^{\nu-1} d_{\sigma_0}(x) < m^\nu \frac{\epsilon}{\mu} = \epsilon,$$

which implies that  $\overline{\sigma_0} \in \Sigma_x$ , a contradiction. Let  $\sigma'_0$  be an hyperbolic element belonging to  $\Sigma_x$ , and note that, by definition,  $\sigma'_0$  is of the form  $\overline{\gamma}$  for some  $\gamma \in \Gamma \setminus \{1\}$ .

By Lemma 2.7(ii) the closed  $\Gamma_x$ -invariant convex set

$$\mathcal{C} = \cap \{\text{Min}(\gamma) : \gamma \in \mathcal{N}_x\}$$

is nonempty. Moreover, since  $\sigma'_0$  is hyperbolic, by Equation (1),

$$\text{Min}(\sigma'_0) = \text{Min}(\sigma_0^m),$$

and since  $\sigma_0^m \in \mathcal{N}_x$ , we derive that  $\mathcal{C} \subset \text{Min}(\sigma'_0)$ . Since  $\sigma'_0 \in \Sigma_x$  we derive that  $\mathcal{C} \subset \tilde{N}$  and since  $x \notin \tilde{N}$  we derive that  $x \notin \mathcal{C}$ . Letting  $c : [0, \infty) \rightarrow X$  be the constant speed geodesic ray with  $c(0) = P_{\mathcal{C}}(x)$  and  $c(1) = x$ , we deduce from Lemma 2.6 that  $\nabla d_\gamma \cdot \dot{c}(1)$  is non-negative for every  $\gamma \in \Gamma_x$  while  $\nabla d_{\sigma'_0} \cdot \dot{c}(1)$  is positive since

$$c(1) = x \notin \tilde{N} \supset \text{Min}(\sigma'_0) \supset \mathcal{C} \ni c(0).$$

Thus, as in Case 1, the directional derivative of  $\tilde{\psi}$  with respect to  $\dot{c}(1)$  is negative, indicating that  $\nabla \tilde{\psi}(x) \neq 0$ .

**Case 3:** Finally, suppose that  $\Gamma_x$  contains a nontrivial parabolic element. By (b),  $\mathcal{N}_x$  contains a nontrivial central parabolic element  $\gamma_0$ . Since  $\gamma_0$  is

central in  $\mathcal{N}_x$  and  $\mathcal{N}_x$  is of finite index in  $\Gamma_x$ , the conjugacy class  $C_{\Gamma_x}(\gamma_0)$  of  $\gamma_0$  in  $\Gamma_x$  is finite. Pick  $\tau$  such that  $d_{\gamma_0}(x) > \tau > \inf(d_{\gamma_0})$  and set

$$\mathcal{C} := \cap_{\gamma \in C(\gamma_0)} \{d_\gamma \leq \tau\}.$$

Then  $\mathcal{C}$  is nonempty by Lemma 2.7(i), and  $x \notin \mathcal{C}$  by the choice of  $\tau$ . Clearly  $\mathcal{C}$  is closed convex and  $\Gamma_x$  invariant. Thus, setting  $\hat{n}$  to be the unit tangent at  $x$  to the ray  $c(t)$  coming from  $P_{\mathcal{C}}(x)$ , we derive from Lemma 2.6 that  $\hat{n} \cdot \nabla d_\gamma(x) \geq 0$  for every  $\gamma \in \Gamma_x$ . We claim that there must be  $\sigma \in \Sigma_x$  for which  $\hat{n} \cdot \nabla d_\sigma(x) > 0$ . Indeed, assuming this is not the case, one derives that for  $\sigma \in \Sigma_x$ ,  $d_\sigma(c(t))$  is locally constant, and since  $X$  is analytic it is globally constant. Hence  $\gamma \cdot c(t)$  is parallel to  $c(t)$  for every  $\gamma \in \Sigma_x$ , and since  $\Sigma_x$  generates  $\Gamma_x$  and parallelity is a transitive relation on geodesics in  $X$ , the same holds for every  $\gamma \in \Gamma_x$ . This however fails to hold for  $\gamma = \gamma_0$ . As in the previous cases, we derive that  $\nabla \tilde{\psi}(x) \cdot \hat{n} < 0$ , and in particular  $\nabla \tilde{\psi}(x) \neq 0$ .  $\square$

## 2.5. Consequences of the main proposition.

**Theorem 2.9.** *For every  $\delta > 0$  there is a deformation retract  $r_\delta : M \setminus N \rightarrow \psi_{\leq \delta}$ .*

*Proof.* By Proposition 2.5 the smooth function  $\psi(x)$  has no positive critical values. We claim furthermore that  $\psi : M \setminus N \rightarrow \mathbb{R}^{\geq 0}$  is a proper map. Indeed, for  $x \in X \setminus \tilde{N}$ ,  $\tilde{\psi}(x) \leq a$  implies that  $\phi_{\bar{\gamma}}(x) \leq a$  for every  $\gamma \in \Gamma \setminus \{1\}$  which implies that  $d_\gamma(x) \geq f^{-1}(a)$  if  $\gamma$  is parabolic, and that  $d_\gamma(x) \geq \frac{f^{-1}(a)}{\mu}$  if  $\gamma$  is elliptic or hyperbolic. Thus the injectivity radius at  $\pi(x)$  is at least  $\frac{f^{-1}(a)}{2\mu}$ . It follows that  $\psi^{-1}([0, a])$  is contained in the  $(\frac{f^{-1}(a)}{2\mu})$ -thick part of  $M$ , which is compact since  $M$  has finite volume.

The result follows by standard Morse theory (c.f. [M], Theorem 3.1).  $\square$

An immediate consequence of Theorem 2.9 is that  $\psi_{\leq \delta}$  is nonempty for every  $\delta > 0$ . The proof of 2.9 also shows that  $\psi_{\leq \delta}$  is compact. Since

$$\tilde{\psi}_{\leq 0} = \bigcap_{\delta > 0} \tilde{\psi}_{\leq \delta}.$$

we derive that the same hold for  $\delta = 0$ :

**Corollary 2.10.** *For every  $\delta \geq 0$ ,  $\psi_{\leq \delta}$  is nonempty and compact.*

Moreover, when  $X$  is not  $\mathcal{H}^2$  we get:

**Corollary 2.11.** *Assume that  $\dim X > 2$ . Then  $\psi_{\leq \delta}$  is connected for every  $\delta \geq 0$ .*

*Proof.* Since  $\Gamma \leq G$  is irreducible and  $G$  is center-free, every  $\gamma \in \Gamma \setminus \{1\}$  projects nontrivially to every factor of  $G$  (cf. [R, Corollary 5.21]). By Corollary 2.2  $\text{codim}_M N \geq 2$  and hence  $M \setminus N$  is connected. Thus by Theorem 2.9  $\psi_{\leq \delta}$  is connected for  $\delta > 0$ , and since a descending intersection of compact connected sets is connected, the same holds for  $\delta = 0$ .  $\square$

**Remark 2.12.** The nonemptiness of  $\psi_{\leq 0}$  implies the stronger version of the Kazhdan–Margulis theorem, since the injectivity radius at points of  $\psi_{\leq 0}$  is at least  $\frac{\epsilon}{2\mu}$ ; The pre-image in  $G$  of the  $\frac{\epsilon}{2\mu}$ -neighborhood of  $K$  in  $G/K$  is an open set which trivially intersect a conjugate of every lattice. Furthermore this set contains a maximal compact subgroup of  $G$ .

**2.6. Finishing the proof.** Assume that  $G$  is not locally isomorphic to  $\text{PSL}_2(\mathbb{R})$ . For  $G = \text{PSL}_2(\mathbb{R})$  the theorem follows from the Gauss–Bonnet formula and the explicate presentation of surface groups.

Since  $\tilde{\psi}_{\leq \delta}$  is connected nonempty and  $\Gamma$  acts freely on it (as all the fixed points of nontrivial elements are in the singular submanifold  $\tilde{N}$ ), we conclude, from standard covering theory, that:

**Corollary 2.13.** *For every  $\delta \geq 0$ ,  $\Gamma$  is a quotient of  $\pi_1(\psi_{\leq \delta})$  — the fundamental group of  $\psi_{\leq \delta} = \pi(\tilde{\psi}_{\leq \delta}) = \Gamma \backslash \tilde{\psi}_{\leq \delta}$ .*

Instead of rescaling the Haar measure of  $G$ , let us suppose that it is fixed and corresponds to the Riemannian measure of  $X$ , and prove that there is a constant  $C(G)$  such that  $d(\Gamma) \leq C(G)\text{vol}(G/\Gamma)$ . Recall from the proof of Theorem 2.9 that  $\psi_{\leq 0}$  is contained in the  $\alpha$ -thick part of  $M$  for  $\alpha = \frac{\epsilon}{2\mu}$ . Let  $\mathcal{S}$  be a maximal  $\alpha$ -discrete subset of  $\psi_{\leq 0}$ . For  $t > 0$  denote by  $v(t)$  the volume of a  $t$ -ball in  $X$ . Since the  $\alpha/2$  balls centered at points of  $\mathcal{S}$  are pairwise disjoint and isometric to a  $\alpha/2$  ball in  $X$ , the size of  $\mathcal{S}$  is bounded by  $\text{vol}(M)/v(\frac{\alpha}{2})$ . Moreover, the union of the  $\alpha$  balls centered at points of  $\mathcal{S}$  covers  $\psi_{\leq 0}$ . Denote this union by  $U$ . Choose  $\delta_0 > 0$  sufficiently small so that  $\psi_{\leq \delta_0}$  is contained in  $U$ .

**Lemma 2.14.**  $\pi_1(\psi_{\leq \delta_0})$  is a quotient of  $\pi_1(U)$ .

*Proof.* The inclusion  $i : \psi_{\leq \delta_0} \rightarrow U$  induces a map  $i_* : \pi_1(\psi_{\leq \delta_0}) \rightarrow \pi_1(U)$ , and the deformation retract  $r = r_{\delta_0}$  restricted to  $U$ ,  $r : U \rightarrow \psi_{\leq \delta_0}$  induces a map  $r_* : \pi_1(U) \rightarrow \pi_1(\psi_{\leq \delta_0})$ . Since  $r \circ i$  is the identity on  $\psi_{\leq \delta_0}$ , we see that  $r_* \circ i_*$  is the identity on  $\pi_1(\psi_{\leq \delta_0})$ . It follows that  $r_* : \pi_1(U) \rightarrow \pi_1(\psi_{\leq \delta_0})$  is onto.  $\square$

Since the sectional curvature is nonpositive on  $M$  the  $\alpha$ -balls centered at points of  $\mathcal{S}$  are convex, and hence any nonempty intersection of such is convex, hence contractible. Thus these balls form a good cover of  $U$ , in the sense of [BT], and the nerve  $\mathfrak{N}$  of this cover is homotopic to  $U$ . Now

---

From here the proof continues as in the special rank one case considered in [BGLS].

$\pi_1(U) \cong \pi_1(\mathfrak{N})$  has a generating set of size  $E(\mathfrak{N})$  — the number of edges of the 1-skeleton  $\mathfrak{N}^1$ . To see this one may choose a spanning tree  $\mathcal{T}$  to the graph  $\mathfrak{N}^1$  and pick one generator for each edge belonging to  $\mathfrak{N}^1 \setminus \mathcal{T}$ . Finally note that the edges of  $\mathfrak{N}$  correspond to pairs of points in  $\mathcal{S}$  which are of distance at most  $2\alpha$ . Thus the degree of each vertex of  $\mathfrak{N}$  is at most  $v(2.5\alpha)/v(\alpha/2)$ . Thus

$$d(\Gamma) \leq |E(\mathfrak{N})| \leq \frac{\text{vol}(M) \cdot v(2.5\alpha)}{2v(\alpha/2)^2},$$

and we can take  $C(G) = \frac{v(2.5\alpha)}{2v(\alpha/2)^2}$ .  $\square$

**Remark 2.15.** The same proof with minor changes applies in the more general context of analytic Hadamard spaces with bounded curvature, as long as the analog of Lemma 2.1 is satisfied: *For every  $d$  there is a constant  $C$  such that for any complete  $d$ -dimensional analytic Riemannian manifold  $M$  with sectional curvature bounded between 0 and  $-1$ , we have*

$$d(\pi_1(M)) \leq C \text{vol}(M),$$

provided the universal cover  $\tilde{M}$  does not admit a submanifold of co-dimension  $\leq 1$  with a nontrivial Euclidean de-Rahm factor

**Acknowledgment:** *The author acknowledges the financial support from the European Research Council (ERC)/ grant agreement 203418.*

**Email:** gelander@math.huji.ac.il

**Address:** Einstein Institute of Mathematics, The Hebrew University, Jerusalem, 91904, Israel

## REFERENCES

- [BGS] W. Ballmann, M. Gromov, V. Schroeder, *Manifolds of Nonpositive Curvature*, Birkhauser, 1985.
- [Be] M. Belolipetsky, Counting maximal arithmetic subgroups. With an appendix by Jordan Ellenberg and Akshay Venkatesh. *Duke Math. J.* 140 (2007), no. 1, 1–33.
- [BGLS] M. Belolipetsky, T. Gelander, A. Lubotzky, A. Shalev, Counting Arithmetic Lattices and Surfaces, to appear in *Ann. of Math.*
- [BL] M. Belolipetsky, A. Lubotzky, Counting manifolds and class field towers, preprint.
- [BT] R. Bott, L.W. Tu, *Differential Forms in Algebraic Topology*, Springer-Verlag, 1982.
- [BH] R. Bridson, A. Haefliger, *Metric Spaces of Non-Positive Curvature*, Springer, 1999.
- [BGLM] M. Burger, T. Gelander, A. Lubotzky, S. Mozes, Counting hyperbolic manifolds, *Geom. Funct. Anal.* 12 (2002), no. 6, 1161–1173.
- [FGT] T. Finis, F. Grunewald, P. Tirao, The cohomology of lattices in  $\text{SL}(2, \mathbb{C})$ , preprint.

- [GR] H. Garland, M.S. Raghunathan, Fundamental domains for lattices in (R-)rank 1 Lie groups, *Ann. of Math.* **92** (1970), 279-326.
- [Ge] T. Gelander, Homotopy type and volume of locally symmetric manifolds. *Duke Math. J.* 124 (2004), no. 3, 459–515.
- [Lu1] A. Lubotzky, Free quotients and the first Betti number of some hyperbolic manifolds. *Transformation Groups* 1 (1996), 71–82
- [Lu2] A. Lubotzky, *On finite index subgroups of linear groups* *Bull. London Math. Soc.* 19 (1987), no. 4, 325–328.
- [K] D.A. Kazhdan, Connection of the dual space of a group with the structure of its closed subgroups, *Functional Analysis and Application* **1** (1967), 63-65.
- [KM] D.A. Kazhdan, G.A. Margulis, A proof of Selberg’s hypothesis. *Math Sbornik* (N.S.) 75 (117), 162-168 (1968) [Russian].
- [LS] A. Lubotzky and D. Segal, *Subgroup Growth*, Birkhauser, 2003.
- [M] J. Milnor, *Morse theory*, Based on lecture notes by M. Spivak and R. Wells. *Annals of Mathematics Studies*, No. 51 Princeton University Press, Princeton, N.J. 1963 vi+153 pp.
- [R] M.S. Raghunathan, *Discrete Subgroups of Lie Groups*, Springer, New York, 1972.
- [SV] R. Sharma and N. Venkataramana, Generators for arithmetic groups, preprint.
- [Th] W.P. Thurston, *Three-Dimensional Geometry and Topology*, Volume 1, Princeton univ. press, 1997.